

Discovering Ontological Correspondences through Dialogue

Gabrielle Santos, Terry R. Payne, Valentina Tamma, and Floriana Grasso

Department of Computer Science, University of Liverpool, L69 3BX
{G.Santos, T.R.Payne, V.Tamma, floriana}@liverpool.ac.uk

Abstract. Whilst significant attention has been given to centralised approaches for aligning full ontologies, limited attention has been given to the problem of aligning partially exposed ontologies in a decentralised setting. Traditional ontology alignment techniques rely on the full disclosure of the ontological models that find the “best” set of correspondences that map entities from one ontology to another. However, within open and opportunistic environments, such approaches may not always be pragmatic or even acceptable (due to privacy concerns). We present a novel dialogue based negotiation mechanism that supports the strategic agreement over correspondences between agents with limited or no prior knowledge of their opponent’s ontology. This mechanism allows both agents to reach a mutual agreement over an alignment through the selective disclosure of their ontological model, and facilitates rational choices on the grounds of their ontological knowledge and their specific strategies. We formally introduce the dialogue mechanism, and discuss its behaviour, properties and outcomes.

1 Introduction

The emergence of annotated data and sophisticated mechanisms for representing formal data models has promoted the proliferation of novel services and systems. These independent services usually commit to their own knowledge model (*ontologies*) and interoperate in an opportunistic fashion in order to perform some task. However, as data models differ, the extent to which the messages are understood can be restricted; thus approaches are necessary to support semantic reconciliation and thus enable seamless interactions to take place between these services. Usually these approaches rely on reaching some form of agreement on the choice of mappings or *correspondences* to translate between the entities in two ontologies. Whilst the problem of determining the vocabulary to use when integrating heterogeneous knowledge has been investigated by numerous research efforts [2, 6, 25], they typically require that both ontological models are shared with some party responsible for discovering the correspondences, even though there may be no guarantee that such correspondences exist; thus this is a limiting assumption. Furthermore, privacy has become increasingly pertinent, whereby neither agent is necessarily prepared to disclose its *full* ontology [11, 15], *e.g.*, if the knowledge encoded within an ontology is confidential or commercially sensitive.

In this paper, we recast this problem as a form of decentralised negotiation, by exploring how dialogue protocols can be used to determine mappings that satisfy each of the agents requirements and strategies. The use of dialogical models allow the agents

to state their position regarding the correctness of some mapping in an asynchronous and distributed fashion, whilst maintaining control over the type of knowledge (class labels vs. ontological model) disclosed. We investigate the issue of reaching an agreement that facilitates the translation of one term from a vocabulary into a corresponding one in a different vocabulary. These translations are not precomputed before any interaction mechanism can be defined, but are rather computed opportunistically (*anytime*) and satisfy the agents requirements and strategies whilst limiting the information exchange only to what is pertinent to support a specific translation. Our main contribution is a dialogue based negotiation mechanism that allows the agents to propose viable lexical mappings and then support these proposals with evidence in the form of ontological fragments, thus collaboratively generating a mutually acceptable partial alignment. These are shared on a per-need basis, and hence the mechanism is purely opportunistic.

This paper is organised as follows: Section 2 introduces the challenge of reconciling heterogeneous knowledge sources, and introduces various constraints that characterise our approach. Section 3 introduces the formalism used throughout the paper, and presents the dialogue protocol and use of arguments to support candidate correspondences. The approach is then illustrated through a walkthrough example in Section 4, and its theoretical properties are discussed in Section 5, before concluding in Section 6.

2 Background & Related Work

The ability to reconcile independently developed knowledge sources is crucial in supporting critical decision making in intelligent applications that require the interaction between disparate knowledge sources. *Ontologies* are machine readable specifications of a conceptualisation of some given domain knowledge [13]; they define the entities and the relationships between them that model such knowledge. It is often the case, however, that the agents differ in the vocabularies (ontologies) they assume, thus compromising seamless *semantic interoperability* between dynamic and evolving systems.

Ontology alignment [7] (the creation of sets of mappings between corresponding entities within a pair of ontologies) can support semantic interoperability between knowledge bases, and thus is an essential component for agent communication. However, even in similar domains, the ontologies can be modelled differently using a variety of modelling languages and contrasting assumptions, which can make translating one ontology into another increasingly difficult. For two systems to accurately and successfully communicate, this semantic heterogeneity between ontologies needs to be resolved.

The ontology alignment community has proposed diverse approaches that *align* ontologies in order to find sets of correspondences; however, alignment approaches are typically centralised processes that require full access to both ontologies. Such approaches try to maximise the number of correspondences created (*coverage*) given some objective function, but they are task agnostic, *i.e.* they do not guarantee to provide correspondences that support a given task or set of queries. Even if an alignment can be found, this might not actually support the representation of a joint task [21]. Furthermore, the axioms defined in the ontology may represent proprietary or commercially sensitive knowledge, and an agent may find it strategically important to impose some restriction over access to this knowledge [11, 15]. Thus, there is a need for alignment

approaches that only generate mappings for the knowledge that is pertinent to some joint task. By structuring the alignment process as a decentralised dialogue, agents can independently determine what axioms they need to expose.

The dialogue based alignment mechanism proposed here is based on the notion of *conversations* as social constructs, where utterances are exchanged in order to achieve some *joint activity* or task [4]; and on the cognitive mechanisms for communication and coordination of activities [14, 22]. The dialogue determines whether there is a *common ground* [5] for establishing the alignment, by generating and sharing justifications for each correspondence proposed. An underlying assumption of our dialogue based approach is that it satisfies the principle of least collaborative effort, where participants try to minimise the total effort spent on a conversation, as typically the fewer exchanges required to clarify references, the better this common ground. It also obeys Grice’s *Co-operative Principle* [12] by assuming that: i) the participating agents are truthful; ii) they make informative contributions as required; and iii) they keep their interactions terse and do not provide more information than necessary. This principle supports the pertinent sharing of knowledge computed on a per-need basis; and further specification is only applied when the communication becomes ineffective. Previous investigations into *meaning negotiation* have also built upon this principle, whereby ontological reconciliation should be *rational*. This problem was first introduced in [2], where *ontology negotiation* was facilitated through a communication protocol that allowed agents to exchange ontological fragments by successively specifying the meaning of given entities. Other studies have addressed different aspects of ontology negotiation [6, 15, 18]. Anemone [6] advocated a lazy, minimal protocol whereby agents exchanged logical definitions in an attempt to define a minimal shared ontology with no information loss. However, it assumed that agents had perfect knowledge over the instances of their ontological models (*i.e.* the underlying approach was grounded through an extensional model), which was used to induce a class description covering certain instances.

Other approaches align heterogeneous ontologies through decentralised negotiation mechanisms [15, 21] or argumentation [17, 18]. In [15], agents selectively exchange details of a priori privately known correspondences, and propose repairs to address any emergent conservativity violations [24], resulting in alignments that are mutually acceptable to both agents without disclosing the full ontological model. Argumentation was used to rationally select correspondences based on the notion of partial-order preferences over their different properties (e.g. structural vs terminological) [18]. This form of correspondence negotiation utilises a course-grained decision mechanism that fails to assess whether or not a correspondence is *acceptable* to each agent (given other mutually accepted correspondences), and assumes that all the correspondences are shared.

3 The Dialogue Mechanism

The dialogue mechanism allows two agents, a *proponent* (a_1) and an *opponent* (a_2), to take turns in exchanging information (through a sequence of *dialogue moves* listed in Table 1) to support a candidate correspondence between the *entities* in their respective ontologies. We assume that an *ontology* \mathcal{O} is modelled as a set of axioms describing these entities, which consist of *classes* N_C and their *relations* N_R . As each agent commits

to its own ontology \mathcal{O} (i.e. agent a_i commits to \mathcal{O}^{a_i}), the entities may be *disclosable* or *private*, depending on the strategy or context of the agent. Thus, the aims of the dialogue are to establish an *alignment* (consisting of a set of *correspondences* [7]) for the entities that are *disclosable* (and thus avoid negotiation over any private entity), whereby each agent negotiates over entities in a *disclosable* signature $\Sigma^d = N_C^d \cup N_R^d$ i.e. the set of disclosable class and property names used in \mathcal{O} . For clarity, we use Σ in the remainder of this paper to refer to the disclosable signature of an agent.

We assume that the ontologies are represented as an edge-labelled directed graph¹ G , where G is an ordered pair $G = (V, E)$ such that:

- $V \subseteq N_C \cup L$ is a finite set of vertices (where L is the set of literals);
- $E \subseteq V \times N_R \times V$ is a ternary relation describing the edges (including labels). As the direction of the edge $e \in E$ represents the ‘subsumes’ relation (\sqsubseteq), two edges are required to represent ‘disjoint’ (\perp) and ‘equivalent’ (\equiv).

We denote with $\pi = \langle s, p, o \rangle$ a subgraph of G (also known as a *triple*) where the disclosable *subject* $s \in N_C^d$ and the disclosable *object* $o \in N_C^d \cup L$ are vertices, and the disclosable *predicate* $p \in N_R^d$ is an edge that relates s to o . We use Π to denote the set of all π .

For two agents to interoperate in an encounter, they need to *align* [7] their respective disclosable vocabulary fragments Σ^{a_1} and Σ^{a_2} , such that the resulting alignment establishes a logical relationship between the disclosable entities belonging to each of the two ontologies. Hence, a correspondence is a mapping between an entity in a source signature (Σ^{a_1}), and a corresponding entity in a target signature (Σ^{a_2}).

Definition 1: A **correspondence** is a triple denoted $c = \langle e, e', r \rangle$ such that $e \in \Sigma^{a_1}$, $e' \in \Sigma^{a_2}$, $r \in \{\equiv, \sqsubseteq, \supseteq, \perp\}$.

We focus on finding concept correspondences, and hence only consider aligning disclosable concept names in $N_C^{a_1}$ and $N_C^{a_2}$. Furthermore, we only consider *logical equivalence* (as opposed to *subsumption* (\sqsubseteq) and *disjointness* (\perp)).²

3.1 Dialogue Protocol

The dialogue protocol comprises a sequence of communicative acts, or *moves* (denoted \mathcal{M}), whereby two participating agents take turn to share statements supporting or refuting a candidate correspondence. For every dialogue move, we assume that each agent plays a role; i.e. a_1 is either a *sender* x or *recipient* \hat{x} (and conversely, a_2 plays the alternate role, such that they never play the same role concurrently). After each move, the agents swap roles, and thus take turns in acting as sender or recipient. The set of legal *moves*, \mathcal{T} , are summarised in Table 1, and their use is illustrated in the walkthrough in Section 4. The syntax of each move is of the form $m = \langle x, \tau, e, e', l \rangle$, where τ is the move type such that $\tau \in \mathcal{T}$, and $\mathcal{T} = \{\text{initiate}, \text{propose}, \text{assert}, \text{accept}, \text{reject}, \text{testify}, \text{justify},$

¹ This is common in ontology alignment approaches [7] and allows us to represent the underlying ontological model irrespectively of the ontology language used (e.g. RDF or OWL).

² This assumption does not affect the generality of our approach, and the majority of ontology alignment approaches that align entities only consider equivalence. Extending the dialogue to support the discovery of subsumption relations is the subject of future work.

Table 1. The set \mathcal{T} of legal moves permitted by the dialogue.

Syntax	Description
$\langle x, \textit{initiate}, e, \textit{nil}, \textit{nil} \rangle$	A new source entity e is proposed, with the aim of finding a possible correspondence.
$\langle x, \textit{propose}, e, e', \textit{nil} \rangle$	A new (<i>i.e.</i> not previously disclosed) candidate entity e' is proposed which lexically matches e .
$\langle x, \textit{justify}, e, e', \textit{nil} \rangle$	A new π is requested to support the candidate correspondence between e and e' .
$\langle x, \textit{testify}, e, e', \pi \rangle$	If an undisclosed π is known that supports the candidate correspondence (with the highest ranking predicate), then it is shared; otherwise $\pi = \textit{nil}$.
$\langle x, \textit{assert}, e, e', A \rangle$	The candidacy of a correspondence between e and e' is asserted, with the supporting argument A containing a subset of disclosed π pairs whose aggregate <i>neighbourhood similarity</i> σ_n supports the candidacy. Note that A and σ_n are presented in Section 3.3.
$\langle x, \textit{accept}, e, e', A \rangle$	The candidacy is accepted if the neighbourhood similarity σ_n of the premise in A is above threshold given the sending agent's similarity metrics.
$\langle x, \textit{reject}, e, e', \textit{nil} \rangle$	The candidacy is rejected if the neighbourhood similarity σ_n of the premise in A is below threshold given the sending agent's own similarity metrics, and no other supporting evidence is available.
$\langle x, \textit{fail}, e, \textit{nil}, \textit{nil} \rangle$	No further undisclosed candidate entities could be found that lexically match e .
$\langle x, \textit{end}, \textit{nil}, \textit{nil}, \textit{nil} \rangle$	The proponent terminates the dialogue.

$\textit{fail}, \textit{end}\}$; e represents the source entity being discussed (identified within the *initiate* move); e' is the current candidate target entity (*i.e.* the entity that could be mapped to from e); and l represents a list of zero or more additional elements (depending on the type of move). For some moves, it may not be necessary to specify the source entity, the target entity or any additional elements, in which case they will be empty or unspecified (represented with *nil*). Fig. 1 illustrates the different states that can occur during the dialogue, and identifies what moves can be legally taken by which agent. The choice of move is determined by the agent's individual strategy (discussed below).

Both agents manage a public knowledge base, or *Commitment Store*³ CS , which contains a trace of all of the moves uttered by each agent [26]. Each agent manages its own private knowledge base, known as the *Gamma Store*⁴ (Γ), that stores private knowledge regarding the ontological structure of the *opponent* that has been garnered through the assertions made in the dialogue. Each of the Gamma Stores contains a partially connected graph, that is:

- either an independent vertex $v_i \in N_C$ representing a candidate concept from the opponent's ontology for inclusion in a correspondence;
- or the *neighbourhood* of the concept v_i , *i.e.* the subgraph originating from the vertex v_i constructed through the exchange of triples that form a directed path from v_i to support its candidacy.

³ Although the agents maintain individual copies of the CS , these will always be identical, and thus we do not distinguish between them.

⁴ We distinguish between the sender's Gamma Store, Γ^x , and the recipient's store, $\Gamma^{\hat{x}}$.

Definition 2: The **lexical similarity metric** is the function $\sigma_l : N_C \times N_C \rightarrow [0, 1]$ which returns the lexical similarity between the labels of two entity names $e, e' \in N_C$, such that $\sigma_l(e, e') = 1$ iff $e = e'$ and 0 if the two labels are different.

This function is used in the initial part of the dialogue to discover those entities in agent a_2 's signature that could lexically match an entity in agent a_1 's signature (*anchors*). A lexical match is considered *viable* if $\sigma_l(e, e')$ is greater or equal to its threshold ϵ_l .

An important component of the dialogue is how the agents share structural details about the ontology in the neighbourhood of an entity under consideration. For any given entity $e \in \Sigma$, there will be a directed path⁶ within the graph that relates e to other entities in its neighbourhood, where the maximum length of the path is bounded by the depth of the ontology. Thus, any triple (π) within this path could be disclosed (*i.e.* shared with the other agent) to provide more details of the entities' local neighbourhood, provided that it forms a path from the entity itself. Depending on the strategy that an agent may adopt, it may assume a depth-first traversal as opposed to breadth-first when disclosing its triples. Therefore, we assume that each agent utilises a function $\text{rank}(e)$ that generates a strict pre-ordering of triples for a given subject e . This is formally defined as:

Definition 3: The **rank function**, $\text{rank} : N_C \rightarrow \mathcal{R} \subseteq \Pi$ returns an ordered list of triples in a path starting at some entity $e \in N_C$, where $\forall \pi_i, \pi_j \in \mathcal{R} : \pi_i \succ \pi_j$.

An agent can request triples belonging to the local neighbourhood of some entity e' in the other agent's ontology, to support the candidacy of a correspondence. We make no assumptions about how the ranking function is defined by any specific agent, and thus the order in which the triples are ranked. Furthermore, following the *Similarity Flooding* approach [19], we also restrict our attention to paths of length 1, and thus only disclose those triples for which e' is a subject.

As *subject-predicate-object* triples relating to e' are disclosed by one agent, the second agent should try to identify similar localised structures in its own ontology. This may be based purely on the triples themselves, or may also take into account other information that has so far been ascertained or inferred. As with the σ_l function, we make no assumptions about how the similarity function is defined, but simply that there is some function for each agent defined formally as:

Definition 4: The **structural similarity metric** is the function $\sigma_s : \Pi \times \Pi \rightarrow [0, 1]$ that returns the structural similarity between two triples $\pi, \pi' \in \Pi$, such that $\sigma_s(\pi, \pi') = 1$ if the two triples are considered as equivalent, and 0 otherwise.

3.3 Arguments and Neighbourhood Similarity

The dialogue mechanism utilises arguments that allow the agents to propose candidate correspondences (between the entities in their respective ontologies), and to justify them or refute them on the grounds of some evidential fact. The agents are assumed to be truthful and to cooperate in order to reach an agreement on the best correspondence to use to map two entities from their respective ontologies.

For this reason, agents can only make arguments that assert the validity of a new correspondence that was not previously disclosed, or question its correctness by stating

⁶ Given the example in Fig. 3, the neighbourhood of $e = \text{Author}$ would include the triple $\langle \text{Author}, \text{hasInitials}, \text{Initials} \rangle$, but would *not* include the triple $\langle \text{Paper}, \text{hasAuthor}, \text{Author} \rangle$.

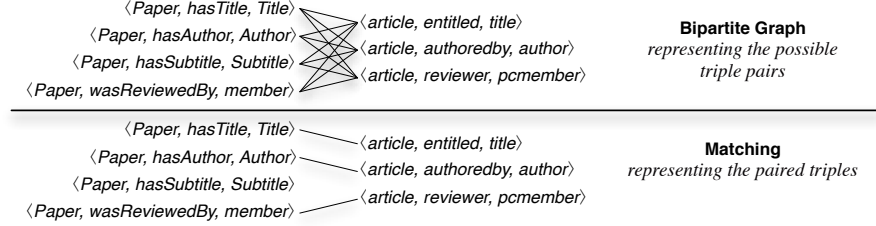


Fig. 2. Possible pairs of triples (top) and a matching (bottom) from the example (Sec. 4).

an alternative correspondence for one of the same entities. As each new argument either introduces a new correspondence, or states a new premise for an existing one, there is no possibility of cycles in arguments, and thus the agents will either reach an agreement or they will reject the proposal. The arguments are defined over the language \mathcal{L} , with the same syntactic primitives as defined for the dialogue. Each agent can form arguments about a candidate correspondence c and entities e in the disclosable signature $\Sigma_d^{a_i}$ of their ontology. \mathcal{L} is the set of formulae ℓ defined by:

$$\ell ::= e|c|(\{\pi\}, c)$$

Hence \mathcal{L} will contain statements about $\Sigma_d^{a_1}$, $\Sigma_d^{a_2}$ and the correspondences mapping from one signature into the other.

Definition 5: An **Argument** is a pair $A = (Pr, Cl)$, where $Pr \subseteq \mathcal{L} \cup \{\top\}$ and $Cl \in \mathcal{L}$. We define $\text{Args}(\mathcal{L})$ the set of all arguments derivable from the language \mathcal{L} .

In this definition, Pr is the *support* (representing a set of premises of an argument), whilst Cl is the *claim*. Facts (*i.e.* statements with no premises) are represented as (\top, Cl) . An argument expresses a relationship between the *claim* and the *support*, such that if the support holds, then the claim must also hold. In our dialogue, the support expresses a *justification* for some neighbourhood similarity (based on a set of related triples) for two entities e and e' , and the claim asserts the viability of a correspondence between these two entities, *i.e.* that the correspondence has some evidence of correctness. The support is based upon some *injective matching* between a bipartite graph (Fig. 2 bottom) representing the triples in an agent's own ontology, and those disclosed by the other agent as part of the dialogue, resulting in matched pairs (π, π') . Each π disclosed by one agent will have some similarity to zero or more triples disclosed by its opponent, as illustrated by the example in Fig. 2 (top) between two example sets of triples supporting a correspondence between the entities *Paper* and *article*.

The *neighbourhood similarity* σ_n is computed over the set of all matching (π, π') pairs (that form a bipartite graph - Fig. 2, bottom), such that no triple from one ontology is "paired" to more than one triple in the other ontology (*i.e.* finding an *injective*, or *one-to-one* mapping between the sets of triples). Depending on the choice of objective function used [9, 16], this can be achieved by finding a *matching* in the graph.

Definition 6: The **neighbourhood similarity** is the function $\sigma_n : \{(\pi, \pi') \in \Pi \times \Pi \mid \pi \in \Gamma, \pi' \in \mathcal{O}\} \rightarrow [0, 1]$ that returns an aggregate similarity calculated from a matching generated from the weighted Bipartite graph obtained by calculating all possible structural similarities between the triples in an agent's Gamma Store Γ and the triples in the disclosable fragment of the opponent's ontology \mathcal{O}^{a_2} , such that $\sigma_n(\pi, \pi') = 1$ if the neighbourhood is structurally equivalent, and 0 otherwise.

As we make no assumption w.r.t. the objective function used to generate the matching (other than assuming that a structural similarity metric σ_s is used to generate the similarity of each pair), we define the function $\text{pairing} : \Pi \times \mathcal{O} \rightarrow \Pi$ that generates a set of triple pairs given the triples in Γ and those in the agents ontology \mathcal{O} .

For example, assuming the triples in Fig 3, the agent *Alice* may have disclosed all four triples to *Bob*. Therefore, *Bob* has:

$$\Gamma^{Bob} = \{ \langle \text{Paper}, \text{hasTitle}, \text{Title} \rangle, \quad \mathcal{O}^{Bob} = \{ \langle \text{article}, \text{reviewer}, \text{pcmember} \rangle, \\ \langle \text{Paper}, \text{hasAuthor}, \text{Author} \rangle, \quad \langle \text{article}, \text{entitled}, \text{title} \rangle, \\ \langle \text{Paper}, \text{hasSubtitle}, \text{Subtitle} \rangle, \quad \langle \text{article}, \text{authoredby}, \text{author} \rangle \} \\ \langle \text{Paper}, \text{wasReviewedBy}, \text{Member} \rangle \}$$

By using the structural similarity metric σ_s , the complete set of possible triple pairs in Fig. 2 (left) can be determined. Assuming some objective function, the matching in Fig. 2 (right) can be generated. Thus, we state that:

$$\text{pairing}(\Gamma^{Bob}, \mathcal{O}^{Bob}) = \{ (\langle \text{Paper}, \text{hasTitle}, \text{Title} \rangle, \langle \text{article}, \text{entitled}, \text{title} \rangle), \\ (\langle \text{Paper}, \text{hasAuthor}, \text{Author} \rangle, \langle \text{article}, \text{authoredby}, \text{author} \rangle), \\ (\langle \text{Paper}, \text{wasReviewedBy}, \text{Member} \rangle, \langle \text{article}, \text{reviewer}, \text{pcmember} \rangle) \}$$

The premise Pr for the claim by agent a_i for some correspondence c will comprise a subset of pairs from the set $\text{pairing}(\Gamma^{a_i}, \mathcal{O}^{a_i})$, with a corresponding aggregate *neighbourhood similarity* σ_n . Although we make no assumption about how σ_n is defined, it could be based on the structural similarity scores σ_s for each triple pair in Pr . A premise Pr is *acceptable* to an agent if $\sigma_n(Pr)$ is greater or equal to a threshold ϵ_n .

4 Walkthrough Example

We illustrate how two agents utilise the dialogue protocol to find an alignment between the public signatures of their ontologies by means of an example. Two agents, *Alice* and *Bob*, each possess a private ontological fragment (Fig. 3). Both agents implement different structural similarity metrics σ_s , and a subset⁷ of the values for different π triple pairs is given in Table 2. For example, the structural similarity⁸ σ_s between the triple $\langle \text{Paper}, \text{hasTitle}, \text{Title} \rangle$ and $\langle \text{article}, \text{entitled}, \text{title} \rangle$ for *Alice*, $\sigma_s^{Alice} = 0.70$, whereas for *Bob* the similarity for this pair is $\sigma_s^{Bob} = 0.68$. In the example dialogue (Table 3), we assume that the dialogue has already commenced, resulting in *Alice* accepting the correspondence $\langle \text{Author}, \text{author}, \equiv \rangle$ in a previous negotiation round (Moves 1-13; the acceptance of this correspondence is illustrated in Move 13 of Table 3). The order in which the dialogue proponent selects entities for exploration is strategic⁹; for this

⁷ Although other similarity pairs have been calculated, these do not appear in the dialogue example (for example, because the distance is lower than those explicitly stated), and thus have not been given for brevity.

⁸ These similarity pairs are not generated a priori, but are calculated during the dialogue.

⁹ As mentioned previously, we do not specify here how the strategic choices are made by each agent, but assume some objective function that determines these choices exists.

example, we assume that the first two entities *Alice* explores are (in order): *Author* and *Paper*. We assume a neighbourhood similarity metric $\sigma_n(Pr)$ calculates the average structural similarity $\bar{\sigma}_s$ of the triple pairs in the premise Pr , with a coefficient that increases asymptotically as the cardinality of Pr increases. The metric is defined as $\sigma_n(Pr) = \bar{\sigma}_s \times (1 - \frac{1}{2(|Pr|+1)})$. We also assume a neighbourhood threshold $\epsilon_n = 0.55$ and a lexical threshold $\epsilon_l = 0.75$.

Move 14: Having previously accepted a correspondence for *Author* (Move 13), *Alice* utters a *initiate* move (state S1 in Fig. 1), to explore a possible correspondence for the next entity from her public signature that she wants to align; which in this case is *Paper*.

Move 15: *Bob* identifies *article* as the most similar entity in his ontology to *Paper* with a lexical similarity $\sigma_l^{Bob}(Paper, article) = 0.82$ (this value is not given in the table). As this is above threshold ϵ_l , he responds with the move $\langle Bob, propose, Paper, article, nil \rangle$.

Move 16: *Alice* now knows that $\langle Paper, article, \equiv \rangle$ is a potential correspondence c (based on *Bob*'s lexical similarity claim). She verifies that her lexical similarity for the entity pair is above threshold (in this case $\sigma_l^{Alice}(Paper, article) = 0.79$). As she is aware that the entity *Paper* has a local neighbourhood (i.e. there is at least one π that has *Paper* as its subject), she asks *Bob* to provide some evidence to justify the candidacy of c . At this point, neither agents have support for c ; i.e. $Pr = \emptyset$.

Move 17: *Bob* (state S4) generates a strict pre-ordering of the properties for *article*, using the function $rank()$; i.e. $rank^{Bob}(article) = \{reviewer, entitled, authoredby\}$. He uses this to determine the next property that has *article* as its *domain* and that has not yet been disclosed (i.e. that has not yet appeared in the commitment store CS). As none of the properties in $rank^{Bob}(article)$ have yet been disclosed, he shares the fact that the highest ranked property *reviewer* relates the two entities *article* and *pcmember*.

Move 18: *Alice* tries to determine if there is sufficient support for c . She realises that $\langle Paper, wasReviewedBy, Member \rangle$ in her ontology is the most similar triple to the one *Bob* disclosed in move 17, with a similarity $\sigma_s^{Alice} = 0.66$ (Table 2). She calculates that the premise $Pr = \{(\langle Paper, wasReviewedBy, Member \rangle, \langle article, reviewer, pcmember \rangle)\}$ has a neighbourhood similarity $\sigma_n^{Alice} = 0.66 \times (1 - \frac{1}{2(|Pr|+1)}) = 0.66 * 0.75 = 0.495$. She will only *assert* an argument for c if this is above the threshold $\epsilon_n = 0.55$. As this is below threshold, she requests additional evidence to justify c .

Move 19: *Bob*'s next highest ranked property that has not been disclosed (i.e. does not appear in CS) whose domain is *article*, is the entity *entitled*. Therefore he shares the triple $\langle article, entitled, title \rangle$.

Move 20: *Alice* checks to see if one of her triples is similar to that disclosed by *Bob* in move 19. Although she has two triples that share their highest similarity with *Bob*'s disclosed triple, she chooses $\langle Paper, hasTitle, Title \rangle$ as the similarity is higher than $\langle Paper, hasSubtitle, Subtitle \rangle$. She adds this to Pr and calculates the neighbourhood sim-

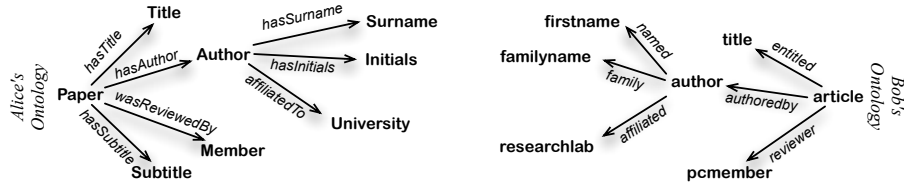


Fig. 3. Two trivial ontology fragments for *Alice* and *Bob* used in the walkthrough example.

Table 2. The structural similarities of possible corresponding triples between *Alice* & *Bob*’s ontologies. Whilst not exhaustive, it lists a subset of triples between the two ontologies.

<i>Alice</i> ’s π	<i>Bob</i> ’s π	σ_s^{Alice}	σ_s^{Bob}
$\langle Author, hasSurname, Surname \rangle$	$\langle author, family, familyname \rangle$	0.76	0.72
$\langle Author, affiliatedTo, University \rangle$	$\langle author, affiliated, researchlab \rangle$	0.85	0.86
$\langle Paper, hasTitle, Title \rangle$	$\langle article, entitled, title \rangle$	0.70	0.68
$\langle Paper, hasAuthor, Author \rangle$	$\langle article, authoredby, author \rangle$	0.65	0.61
$\langle Paper, hasSubtitle, Subtitle \rangle$	$\langle article, entitled, title \rangle$	0.68	0.84
$\langle Paper, wasReviewedBy, Member \rangle$	$\langle article, reviewer, pcmember \rangle$	0.66	0.60

ilarity $\sigma_n^{Alice} = (0.66 + 0.7)/2 \times (1 - \frac{1}{2(2+1)}) = 0.68 * 0.8\dot{3} = 0.56$, which (from *Alice*’s perspective) is above threshold. Therefore she proposes the argument A for the correspondence $c = \langle Paper, article, \equiv \rangle$, given that:

$$Pr = \{(\langle Paper, wasReviewedBy, Member \rangle, \langle article, reviewer, pcmember \rangle), (\langle Paper, hasTitle, Title \rangle, \langle article, entitled, title \rangle)\}$$

Move 21: Given the argument A for the correspondence c asserted in the previous move, *Bob* (state S5) can make one of two possible moves:

- *accept* the argument A if $\sigma_n^{Bob}(Pr)$ is above threshold, and transition to state S1;
- *justify* the candidacy of c by requesting further support (if other undisclosed properties exist).

In this case, *Bob* calculates that the neighbourhood similarity (from his perspective) is $\sigma_n^{Bob} = (0.60 + 0.68)/2 \times (1 - \frac{1}{2(2+1)}) = 0.64 * 0.8\dot{3} = 0.53$, which is below threshold. However, *Bob* is aware of other triples for the entity *article* that do not appear in Pr , and thus asks *Alice* if she could provide some further evidence to justify c .

Move 22: *Alice* now generates her own strict pre-ordering of the properties for *Paper*, using the function $\text{rank}()$; i.e. $\text{rank}^{Alice}(Paper) = \{hasTitle, hasAuthor, hasSubtitle, wasReviewedBy\}$. She shares the triple $\langle Paper, hasAuthor, Author \rangle$ as *hasAuthor* is her highest ranked, non-disclosed property for the domain entity *Paper* (property *hasTitle* was ranked higher but was disclosed in her previous *assert* move).

Move 23: *Bob* recalculates the mean similarity for the new support (inclusive of the triple shared by *Alice* in Move 22): $\sigma_n^{Bob} = (0.60 + 0.68 + 0.61)/3 \times (1 - \frac{1}{2(3+1)}) = 0.63 * 0.875 = 0.551$, which is above threshold. *Bob* is happy to accept the candidacy of c . It is now his turn to *assert* the new argument for c given the new premise Pr .

Move 24: *Alice* confirms that from her perspective, $\sigma_n^{Alice} = (0.66 + 0.7 + 0.65)/3 \times (1 - \frac{1}{2(3+1)}) = 0.67 * 0.875 = 0.59$, which is above threshold, and accepts the argument.

At this point, through co-operation, the agents were able to engage in the joint activity of determining a correspondence between two entities based on the similarity of the local neighbourhood of the entities. Although all of *Bob*’s triples were disclosed, *Alice* was able to reach the consensus without revealing knowledge of one of her triples: $\langle Paper, hasSubtitle, Subtitle \rangle$, even though from *Bob*’s perspective, it was actually more similar to *Bob*’s triple $\langle article, entitled, title \rangle$ than $\langle Paper, hasTitle, Title \rangle$. If in move 20, *Alice* had found that the triple with the highest similarity to $\langle article, entitled, title \rangle$ was actually $\langle Paper, hasSubtitle, Subtitle \rangle$, then *Bob* would have accepted the support in move 21 (as $\sigma_n^{Bob} = (0.6 + 0.84)/2 \times (1 - \frac{1}{2(2+1)}) = 0.67 * 0.8\dot{3} = 0.56$, which was above threshold) and fewer properties would have been disclosed.

Table 3. The messages exchanged between *Alice* and *Bob* in the example dialogue fragment (note that the moves 1-12 are not shown for brevity).

Move	Locution
13	$\langle \text{Alice}, \text{accept}, \text{Author}, \text{author}, (\langle \text{author}, \text{family}, \text{familyname} \rangle, \langle \text{Author}, \text{hasSurname}, \text{Surname} \rangle), (\langle \text{author}, \text{affiliated}, \text{researchlab} \rangle, \langle \text{Author}, \text{affiliatedTo}, \text{University} \rangle), \langle \text{Author}, \text{author}, \equiv \rangle) \rangle$
14	$\langle \text{Alice}, \text{initiate}, \text{Paper}, \text{nil}, \text{nil} \rangle$
15	$\langle \text{Bob}, \text{propose}, \text{Paper}, \text{article}, \text{nil} \rangle$
16	$\langle \text{Alice}, \text{justify}, \text{Paper}, \text{article}, \text{nil} \rangle$
17	$\langle \text{Bob}, \text{testify}, \text{Paper}, \text{article}, \langle \text{article}, \text{reviewer}, \text{pcmember} \rangle \rangle$
18	$\langle \text{Alice}, \text{justify}, \text{Paper}, \text{article}, \text{nil} \rangle$
19	$\langle \text{Bob}, \text{testify}, \text{Paper}, \text{article}, \langle \text{article}, \text{entitled}, \text{title} \rangle \rangle$
20	$\langle \text{Alice}, \text{assert}, \text{Paper}, \text{article}, (\{ (\langle \text{Paper}, \text{wasReviewedBy}, \text{Member} \rangle, \langle \text{article}, \text{reviewer}, \text{pcmember} \rangle), (\langle \text{Paper}, \text{hasTitle}, \text{Title} \rangle, \langle \text{article}, \text{entitled}, \text{title} \rangle) \}, \langle \text{Paper}, \text{article}, \equiv \rangle) \rangle$
21	$\langle \text{Bob}, \text{justify}, \text{Paper}, \text{article}, \text{nil} \rangle$
22	$\langle \text{Alice}, \text{testify}, \text{Paper}, \text{article}, \langle \text{Paper}, \text{hasAuthor}, \text{Author} \rangle \rangle$
23	$\langle \text{Bob}, \text{assert}, \text{Paper}, \text{article}, (\{ (\langle \text{article}, \text{reviewer}, \text{pcmember} \rangle, \langle \text{Paper}, \text{wasReviewedBy}, \text{Member} \rangle), (\langle \text{article}, \text{entitled}, \text{title} \rangle, \langle \text{Paper}, \text{hasTitle}, \text{Title} \rangle), (\langle \text{article}, \text{authoredby}, \text{author} \rangle, \langle \text{Paper}, \text{hasAuthor}, \text{Author} \rangle) \}, \langle \text{Paper}, \text{article}, \equiv \rangle) \rangle$
24	$\langle \text{Alice}, \text{accept}, \text{Paper}, \text{article}, (\{ (\langle \text{article}, \text{reviewer}, \text{pcmember} \rangle, \langle \text{Paper}, \text{wasReviewedBy}, \text{Member} \rangle), (\langle \text{article}, \text{entitled}, \text{title} \rangle, \langle \text{Paper}, \text{hasTitle}, \text{Title} \rangle), (\langle \text{article}, \text{authoredby}, \text{author} \rangle, \langle \text{Paper}, \text{hasAuthor}, \text{Author} \rangle) \}, \langle \text{Paper}, \text{article}, \equiv \rangle) \rangle$

5 Dialogue properties

It is customary to analyse dialogue systems in terms of their *soundness*, *completeness* and *termination* properties. Usually these are not considered in isolation, but they are analysed with respect to the compliance shown by the dialogue to the specific agents' strategies. For instance, a *sound* dialogue protocol result can be roughly restated as obtaining a "successful" dialogue results, i.e. verifying that the claim of the dialogue is "acceptable" w.r.t. the adopted strategies [8].

One of the main characteristics of our approach is that the strategy definition is tightly dependent on the specific choices the agents make in terms of similarity measures. Whilst we argue that having a generic framework is a strength of the presented approach as it makes it customisable to suit different interoperability scenarios, this makes it complex to characterise soundness and completeness. Termination is more straightforward to prove as it is independent of the agents' strategies. Regarding soundness it is important to point out that this does not correspond to correctness with respect to a gold standard alignment. Indeed, it is possible to imagine that in a cooperative domain, agents would behave in an intelligent manner in order to be able to influence the outcome of the dialogue and always arrive at the best possible outcome given their internal knowledge and strategies.

The dialogue presented in the previous section allows agents to only put forward new arguments, either by proposing a new correspondence or by providing evidence supporting some candidate correspondence. Once arguments are uttered, they cannot be retracted. The monotonic property of this dialogue helps us to characterise soundness in terms of obtainable outcomes. Indeed, it is possible to clearly identify two possible outcomes of the dialogue, *fail* and *accept*, leading to the dialogue termination at S0, and the state transitions that cause these outcomes to be reached. Either outcomes represent acceptable solutions to the alignment problem, with *fail* explicitly capturing the fact that the agents cannot find a suitable solution within the constraints dictated by their strategies. The conditions underlying these outcomes are described below, by referring to the states in the diagram in Fig. 1. The pathways to failure are described below:

S2: The proponent initiates the dialogue requesting a match for an entity e (S1), however no entity e' in the opponent signature is a viable match for e , i.e. $\forall e' \in N_C^{\hat{x}} \sigma_l(e, e') < \epsilon_l$.

S3: Following S1, the opponent responds with an entity e' . The proponent then evaluates the potential correspondence (e, e') : if this is not viable (i.e. $\sigma_l(e, e') < \epsilon_l$) then it rejects it, and the dialogue fails. If the correspondence is viable then the proponent might still request the opponent to provide further evidence supporting this proposal, and hence enter a *justify-testify* loop (S3-S4). If the evidence provided is not deemed sufficient, the proponent can reject the correspondence.

S7: Following S3, the proponent assesses the correspondence proposed, and on finding it suitable she asserts it (S5). This assertion however requires some verification from the opponent, who requests that the proponent provides some supporting evidence for the assertion through a *justify-testify* loop (S6-S7, but this time with the proponent being the opponent, and vice versa). If the opponent deems that the evidence is not sufficient it will reject the assertion made by the proposer and the dialogue will fail.

The pathways for the successful termination of the dialogue are clearly identifiable:

S3: Following S1, the opponent responds with an entity e' . The proponent then evaluates the potential correspondence (e, e') and finds that it satisfies its strategy, and hence asserts the viability of the correspondence requiring further evidence (S3-S4). However the proponent may also require further evidence from the opponent (*justify-testify* loop, S3-S4). If the evidence is deemed sufficient, then the proponent asserts the acceptability of the correspondence from his side. This is then evaluated by the opponent who can confirm the acceptability of the correspondence with respect to its strategy (S6), and the dialogue terminates successfully.

S5: The opponent might also require further evidence (S6, S7) and if satisfied, it can assert the correspondence as viable from his perspective, and then this is assessed by the proposer (S8) with or without requiring supporting evidence. If the evidence is requested, then it will be assessed and if it is deemed sufficient (according to the agent's internal strategy) then it will be accepted.

Regarding completeness, it is trivial to see that the dialogue is not complete. The dialogue effectively approximates a greedy search over the space of possible correspondences. This approximation is not guaranteed to be complete as solutions will only be accepted by the proponent of an assertion only if it deems the evidence sufficient for the claim. The same assertion will only be accepted by the opponent if it also deems the

evidence sufficient and if not it will request further evidence to be put forward. However, this mechanism allows the agents to find a solution without exploring all candidate solutions, hence it satisfies the minimality requirement following Grice's maxims.

Given that the dialogue admits only two possible outcomes, and that it cannot propose correspondences or supporting evidence already proposed it is trivial to show that the dialogue terminates.

Proposition 1. *The negotiation dialogue with the set of moves \mathcal{M} in Section 3.1 will always terminate.*

Proof. Sketch. Both agents have finite disclosable signatures, and can only propose one entity to align at a time. Once the entity is proposed, the agents can request that the correspondence is justified in terms of its support; however, this support is also finite, being bounded by the size of the disclosable signature of the ontology. At any point in the dialogue, agents can only add new evidence or assert new correspondences (after having rejected a previous proposal), but are prohibited from revisiting either a correspondence or some evidence previously discussed (i.e. agents can only add to the Commitment Store and not retract from it). If the dialogue does not end before every possible viable correspondence is considered (states S1-S3), then it will end, in the worst case, once the (finite) set of *testify* - *justify* moves providing evidence for the correspondence in the claim have all been made. If no appropriate evidence is provided, then the dialogue will terminate following a *fail* outcome. \square

6 Conclusions

We present work on a dialogue based mechanism that allows agents to reach agreement over an alignment between the disclosable entities of their respective ontologies, without the need for prior information of the ontological structures used by either agent, or some centralised machinery. The proponent takes turns to ask questions about a potential correspondence to ascertain if there is sufficient evidence to support it; and the opponent, through introspection accepts, rejects or seeks further or more compelling evidence to support the claim. A dialogue protocol is introduced that allows agents to reach an agreement over mutually acceptable correspondences, and discusses its properties. It is illustrated through an example that shows how the dialogue is used to establish whether two entities in two different ontologies can be mapped, and the formal properties of the dialogue (w.r.t. soundness, completeness and termination) are presented.

References

1. Ankolekar, A., Burstein, M., Hobbs, J., Lassila, O., McDermott, D., Martin, D., McIlraith, S., Narayanan, S., Paolucci, M., Payne, T., Sycara, K.: DAML-S: Web Service Description for the Semantic Web. In: First International Semantic Web Conference (ISWC) Proceedings. pp. 348–363 (2002)
2. Bailin, S., Truszkowski, W.: Ontology negotiation: How agents can really get to know each other. In: Innovative Concepts for Agent-Based Systems, LNCS, vol. 2564, pp. 320–334. Springer (2003)

3. Cheatham, M., Hitzler, P.: String similarity metrics for ontology alignment. In: *The Semantic Web-ISWC 2013*, pp. 294–309. Springer (2013)
4. Clark, H.H.: *Using language*. Cambridge university press (1996)
5. Clark, H.H., Schaefer, E.F.: Contributing to discourse. *Cognitive science* 13(2), 259–294 (1989)
6. van Diggelen, J., Beun, R.J., Dignum, F., van Eijk, R.M., Meyer, J.J.C.: Anemone: an effective minimal ontology negotiation environment. In: *Proc. AAMAS '06*. pp. 899–906 (2006)
7. Euzenat, J., Shvaiko, P.: *Ontology Matching*. Springer-Verlag (2007)
8. Fan, X., Toni, F.: On computing explanations in abstract argumentation. In: *ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014)*. pp. 1005–1006 (2014)
9. Gale, D., Shapley, L.S.: College admissions and the stability of marriage. *The American Mathematical Monthly* 69(1), 9–15 (1962)
10. Grau, B.C., Dragisic, Z., Eckert, K., Euzenat, J., et al.: Results of the ontology alignment evaluation initiative 2013. In: *Proc. 8th ISWC workshop on ontology matching (OM)*. pp. 61–100 (2013)
11. Grau, B.C., Motik, B.: Reasoning over ontologies with hidden content: The import-by-query approach. *Journal of Artificial Intelligence Research (JAIR)* 45, 197–255 (October 2012)
12. Grice, H.P.: Logic and conversation. In: Cole, P., Morgan, J.L. (eds.) *Syntax and Semantics*, vol. 3: *Speech Acts*, pp. 41–58. Academic Press (1975)
13. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum.-Comput. Stud.* 43(5-6), 907–928 (Dec 1995)
14. Hulstijn, J.: *Dialogue models for inquiry and transaction*. University of Twente (2000)
15. Jiménez-Ruiz, E., Payne, T.R., Solimando, A., Tamma, V.: Limiting logical violations in ontology alignment through negotiation. In: *Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning, KR'16* (2016)
16. Kuhn, H.W.: The hungarian method for the assignment problem. *Naval research logistics quarterly* 2(1-2), 83–97 (1955)
17. Laera, L., Blaoe, I., Tamma, V., Payne, T., Euzenat, J., Bench-Capon, T.: Argumentation over ontology correspondences in MAS. In: *Proc. AAMAS '07*. pp. 1285–1292 (2007)
18. Laera, L., Tamma, V., Euzenat, J., Bench-Capon, T., Payne, T.: Reaching agreement over ontology alignments. In: *The Semantic Web-ISWC 2006*, pp. 371–384. Springer (2006)
19. Melnik, S., Garcia-Molina, H., Rahm, E.: Similarity flooding: A versatile graph matching algorithm and its application to schema matching. In: *Proceedings of the 18th International Conference on Data Engineering, San Jose, CA, USA, February 26 - March 1, 2002*. pp. 117–128 (2002), <http://dx.doi.org/10.1109/ICDE.2002.994702>
20. Miller, G.: *WordNet: a lexical database for English*. ACM Press, New York (1995)
21. Payne, T.R., Tamma, V.: Negotiating over ontological correspondences with asymmetric and incomplete knowledge. In: *Proc. AAMAS '14*. pp. 517–524 (2014)
22. Searle, J.R.: *Speech acts: An essay in the philosophy of language*, vol. 626. Cambridge university press (1969)
23. Shvaiko, P., Euzenat, J.: Ontology matching: state of the art and future challenges. *Knowledge and Data Engineering, IEEE Transactions on* 25(1), 158–176 (2013)
24. Solimando, A., Jiménez-Ruiz, E., Guerrini, G.: Detecting and correcting conservativity principle violations in ontology-to-ontology mappings. In: *Int'l Sem. Web Conf. (ISWC)*. pp. 1–16 (2014)
25. Spiliopoulos, V., Vouros, G.A.: Synthesizing ontology alignment methods using the maximum algorithm. *IEEE TKDE* 24(5), 940–951 (2012)
26. Walton, D., Krabbe, E.: *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. SUNY series in Logic and Language, State University of New York Press (1995)